DOCUMENT RESUME

ED 411 825 IR 056 652

AUTHOR	Hildyard, C. J.; Whitaker, B. J.
TITLE	Chemical Publishing on the Internet: Electronic Journals--Who Needs Them?
PUB DATE	1996-00-00
NOTE	10p.; In: Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3-5, 1996); see IR 056 631.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; Electronic Libraries; *Electronic Publishing; Electronics; Foreign Countries; Internet; Online Systems; Programming Languages; Scholarly Journals; *Science Materials; Technological Advancement; User Needs (Information); *World Wide Web; Writing for Publication
IDENTIFIERS	Electronic Media; HTML; Java Programming Language; Multimedia Technology; SGML

ABSTRACT

The rapid growth in the use of the Internet in recent years has in large part been due to the interest in and use of the World Wide Web. A reader can, in principle, locate and access information related to a specific query from servers located anywhere in the world. The ability to embed multimedia files into Web documents (in either HTML or HTF) also offers authors techniques with which they can extend what can be said in words alone. This paper examines the possibilities offered by the World Wide Web and related technologies in the area of scientific publication. Work within the Electronic Libraries Program (eLib) project to establish an electronic version of the Royal Society of Chemistry's journal, "Chemical Communications," is described. The particular focus is on issues related to document preparation from Standard Generalized Markup Language (SGML) master documents, and the development of multimedia methods for the presentation of complex scientific data, such as three-dimensional molecular structural information, using MIME and Java. The value of the Web for scientific and technical publication is clear, but the structure of the Web and current modes of use also raise difficult issues, and in the conclusion of this paper, the impact of electronics on the publishing process itself is addressed. (Contains 19 references.) (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 411 825

Chemical Publishing on the Internet: Electronic Journals – Who Needs Them?

By:

C.J. Hildyard

B.J. Whitaker

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B.P. Jeapes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IR056052

BEST COPY AVAILABLE

Chemical publishing on the Internet: electronic journals — who needs them?

C.J. Hildyard and B.J. Whitaker
School of Chemistry, University of Leeds, UK

Abstract: *The rapid growth in the use of the Internet in recent years has in large part been due to the interest in and use of the World Wide Web. Indeed, the expansion in this form of electronic communication is so striking that it has led some commentators to make comparisons between the development of the Web and the introduction of the printing press in fifteenth century Europe. As a medium for publication and information dissemination, superficially at least, electronics promises much. Using mechanisms built into the structure of the Web, such as hypertext, information sources can be linked across the globe. So called spiders and robots can then 'crawl' the Web to construct indexes and other search tools. Thus a reader can in principle locate and access information related to a specific query from servers located anywhere in the World. The ability to embed multimedia files into Web documents (in either HTML or HTF) also offers authors techniques with which they can extend what can be said in words alone. In this paper we examine the possibilities offered by the World Wide Web and related technologies in the area of scientific publication. Work within the Electronic Libraries Program (eLib) project to establish an electronic version of the Royal Society of Chemistry's journal,* Chemical Communications, *is described. We focus in particular on issues related to document preparation from SGML master documents, and the development of multimedia methods for the presentation of complex scientific data, such as three-dimensional molecular structural information, using MIME and Java. The value of the Web for scientific and technical publication is clear, but the structure of the Web and current modes of use also raise difficult issues, and in the conclusion we attempt to assess critically the impact of electronics on the publishing process itself.*

Keywords: *electronic publishing, chemistry, World Wide Web, markup languages*

1. Introduction

There are many issues to consider when developing an electronic version of a journal, from the perspective of the publisher, the author and the reader alike. The opportunities for all are obvious. From the publishers' point of view, to name but a few, there are opportunities to speed up the process of publication, to create 'value added' products using multimedia and to monitor document access and usage by readers. From the point of view of the author and the reader the attraction lies in the novel ways in which data can be presented and interacted with. More fundamentally, the World Wide Web (WWW) enables access to a wide variety of information on-demand, conveniently and efficiently.

The drawbacks of the electronic medium can in the main be put down to lack of precedent. The technology is relatively new and is developing so rapidly that there has simply not been time to legislate for problems such as copyright, privacy and data protection on the Internet. However, the obvious promise that this medium offers has led to a number of projects experimenting with the possibilities of electronic publication. It is to be hoped that between them, a framework for electronic publication can be worked out which addresses these issues without being unduly restrictive. The following is a brief description of the aims, strategy and experience of one such project.

2. The CLIC consortium

The CLIC consortium is a project within the Electronic Libraries (eLib) programme which aims to establish an electronic version of the Royal Society of Chemistry's flagship journal, *Chemical Communications*, to run in parallel with the printed version (Ref 1). The eLib programme itself was set up by the Joint Information Systems Committee (JISC) following the recommendations of the Follett report (Ref 2) into the use of information technology (IT) in the electronic library. A key motivation is to investigate whether IT can help to alleviate some of the problems of university libraries today—principally the ever spiralling cost of journal subscriptions and the

associated storage costs. Although mainly used by structural and organic chemists *Chem. Comm.* is a general chemistry journal which aims to disseminate rapidly internationally important research results. The journal takes great pride in its speed of publication, averaging 105 days from receipt of manuscript to publication in 1995, and in the high quality of its content and refereeing (rejection rate *ca.* 50%). The journal is published biweekly and each issue typically runs to over 150 printed A4 pages containing some 80 articles. Each issue also contains a longer feature article highlighting the work of a leading chemist.

Whilst developing a strategy for the implementation of an electronic publication, several issues have arisen. Three important areas will be discussed here.

(1) document archiving — the electronic storage of the research papers;

(2) document delivery — the method by which documents are provided to the user;

(3) 'added value', exploiting to the greatest possible extent the features of the WWW, providing information which by its nature could not be conveyed by the printed page.

2.1. Document archival

Papers are submitted to the Royal Society of Chemistry (RSC) publications offices in Cambridge on paper and in electronic formats, such as Microsoft Word, WordPerfect and to a lesser extent TeX and LaTeX. The RSC encourages authors to submit manuscripts on disk or by e-mail, but currently only about 10% of submissions are so received (although about 40% of submissions are accompanied by discs containing graphical information such as figures, etc.). Of these the vast majority (>90%) are in Microsoft Word format. The rapid development of document preparation software, however, ensures that at any given time there are dozens of formats concurrently in use by potential authors. Obviously the transfer of information between computer systems relies on the compatibility of these formats (which can cause headaches for the publisher). More crucially the constant development of document formats posits a fundamental difficulty for the long term archiving of electronic documents due to the possibility, not to say probability, of formats becoming obsolete. A publisher who is naturally responsible for the long term storage as well as the dissemination of such electronic data therefore faces a real problem.

The solution adopted by the RSC, along with many other publishing houses, has been to make use of Standard Generalised Markup Language (SGML) (Refs 3, 4). Markup is the formatting information which tells a word processor or display client how the document should be rendered, for example font type and size, alignment, etc. Sections of text are usually marked up by placing the contents between two delimiters or tags. Users of TeX (Ref 5), in which the markup is inserted manually, will be more familiar with this concept than those who use WYSIWYG software, in which the markup is largely invisible. SGML is an alternative type of markup that is descriptive rather than procedural. It defines *containers* which specify the semantic content of the textual elements rather than their appearance. Each SGML document starts with a *Document Type Definition* (DTD), which lays out the rules for the markup of a document in terms of the elements it contains. This allows the markup to be designed with the specific data types likely to be encountered in mind — a feature which is particularly useful in specialised subject disciplines such as chemistry and one to which we shall return later.

To give an example of descriptive vs. procedural markup, in SGML this paper might be marked up as:

```
<!DOCTYPE article PUBLIC "-//Online/DTD article//EN">
<ARTICLE>
<TITLE>Chemical Publishing on the Internet ... </TITLE>
<ADDRESS>School of Chemistry, University of Leeds ... </ADDRESS>
<AUTHOR>C.J. Hildyard and B.J. Whitaker</AUTHOR>
<ABSTRACT>The rapid growth in the use of the Internet ... </ABSTRACT>
<SECTION>
<SECTIONHEADING>Introduction</SECTIONHEADING>
There are many issues to consider when developing ...
</SECTION>
...
</ARTICLE>
```

as opposed to, say

```
{\rtf1\ansi \deff4\deflang1033{\fonttbl{\f4\froman\fcharset0\fprq2 Times
New Roman;}{\f8\froman\fcharset0\fprq2 Times;}}
{\stylesheet{\f4\lang2057 \snext0 Normal;} ...
\f4\lang2057 {\b\fs28 Chemical Publishing ... \par }
\par \pard C.J. Hildyard and B.J. Whitaker
\par
\par School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.
\par
\par \pard \qj {\b Abstract}
\par \pard \qj \tab The rapid growth in the use of the Internet ...
\par \pard \qj
\par \pard \qj {\b 1.\tab Introduction}
\par \pard \tab There are many issues to consider when developing ...
```

which is the way the document appears marked up in Rich Text Format (RTF). The latter contains information about the font size, character set, etc. and defines the way in which the document should appear on the page in addition to its content.

In SGML, on the other hand, the markup simply defines the type of each textual element. The first line contains a *document type declaration* identifying the DTD required for the document — in this case a (fictitious) declaration for a document in the style of an English language (EN) article for these proceedings. This is followed by a start tag <ARTICLE>, identical to the tag used for the DTD declaration, which is called the base document element and which activates the DTD. The elements following the <ARTICLE> start tag are fairly self-explanatory. Finally the end of the document is identified with a closing end tag </ARTICLE>. Notice that only the logical structure of the document is important, for example the <ADDRESS>...</ADDRESS> container appears out of sequence from that in the printed version of this article. Textual elements may be nested so, for example, we could define a container <CITATION>...</CITATION> with sub elements <Author>...</Author>, <Year>...</Year>, <Pages>...</Pages> and so on. SGML tags may be also qualified by <I>attributes as in <CITATION type=Book> which might make certain sub elements mandatory, e.g. <Publisher>. These might be optional for other attributes, e.g. with <CITATION type=Journal> it would be unusual to specify the publisher. Each SGML document, known as an *SGML document entity*, consists of three components: an SGML declaration defining the syntax and character set used within the document, the DTD defining the logical structure of the document, and the text itself.

Since the markup in SGML defines the logical structure of the document rather than its appearance, the danger of the data becoming obsolete is significantly reduced — the philosophy being that whether a document is rendered on paper or on a CRT its basic structure is immutable. The final appearance of the document is determined by a post-processor. In our work this is done using CoST (Ref 6). CoST is a Tcl-based SGML post-processing tool, built on sgmls, a public domain SGML parser developed by James Clark (Ref 7). Restricting the formatting stage to a post-processor rather than including the information along with the data itself has clear advantages. Specifically in the context of publishing on the WWW it addresses the problem caused by the fact that the 'standard' Web markup, Hyper Text Markup Language (HTML) is anything but a standard. Currently at version 2 HTML (Ref 8) lacks significant functionality when it comes to rendering scientific and non-English language text. For example, the lack of support for Greek characters creates difficulties in properly rendering scientific documents, particularly mathematical formulae.

The solution adopted by most publishers, including ourselves, is to link to a library of GIF images of the characters in question. This is a crude kludge. It works but it has unfortunate side effects. It adds to the overhead on document transfer and significantly affects download times adversely. Furthermore, since font size is both browser and context dependent (e.g. headings are rendered in a larger font than normal text), there is no way of knowing whether the image supplied is the appropriate size. This can lead to uneven line spacing and a poor overall appearance on the client, and from our limited 'market testing' to date is unpopular with readers.

The latest implementation of HTML, HTML3.2 (Ref 9) (at the time of writing still a proposal) does not address the problem of character sets either, although some extra functionality is added. The question of standards is an important one. HTML 3.2 is not a standard as such but a specification drawn up by the World Wide Web Consortium (W3C), with input from various vendors including Netscape Communications Corporation, Microsoft, IBM, Sun Microsystems, Novell, SoftQuad and Spyglass. The likelihood is that standards will be replaced by periodically updated specifications of the current practice in browser technology. Developments are therefore sure to be rapid and publishers must be able to react swiftly. This is an unsatisfactory situation and certainly not one that a major commercial publishing house can easily accept. By maintaining the document in SGML form and subsequently converting it to HTML, changes in the definition and functionality of HTML can be handled by altering a few lines of code in the translator rather than the source data itself. This not only protects the integrity of the database in a recognised ISO format but allows a publisher to react quickly to developments in WWW servers and browsers. The document archive is in a sense 'future proof'.

The DTD used in the CLIC project is a modified form of the ISO standard 12083 DTD which was designed to cater specifically for scientific and technological applications (Ref 10). However, in order to handle the vast array of chemical information and data types, the DTD needs to be extended. Chemistry Markup Language (CML) (Ref 11) is a proposal designed to address this problem by defining semantically rich chemistry-specific markup to represent the diversity of chemical data as completely as possible. It is only with a sufficiently 'tag-rich' DTD that one can properly make use of the variety of possibilities offered by the Web. We expand on this point in Section 2.3.

2.2. Document delivery

Once the SGML archive has been established one must decide on the method of delivering the document to the user. It would be counter-productive to have an SGML database and an HTML database containing the same information in existence at the same time. This also goes against underlying principle of any database, namely that there should only be one master. The approach used in the CLIC project is to generate the HTML 'on-the-fly' using CoST via a *Common Gateway Interface* or CGI script. This is executed as and when the user requests an article, and removes the necessity for a parallel HTML archive. Although conversion times are rapid — 1-2 CPU seconds on a Sparc2 for a typical *Chem. Comm.* paper — the converted articles may be temporarily cached, so that a future reader requesting the same article a short time later does not have to wait for the conversion process to be repeated. If the lifetime of documents held in the cache is sufficiently short, updates to the code will filter through to the user very quickly. Thus there will be only one permanent archive and a dynamic

cache of temporary documents, in the interests of efficiency and speed of delivery.

CoST operates on the Element Structure Information Set (ESIS) which is produced by sgmls and not the SGML document instance itself. This ensures that the SGML is validated against the DTD before being processed, providing an extra safeguard for the integrity of the archived SGML data. SGML documents can be seen in terms of a tree structure which CoST navigates using a powerful query language. The contents of each tag in the SGML document instance are associated with a corresponding HTML tag, with hyperlinks generated where necessary. As long as there is sufficient information within the SGML document instance, or in other words the DTD allows for a comprehensive markup of the document structure, any HTML features or indeed non-HTML external applications can be generated automatically. CoST can also be used to generate the Graphical Abstracts pages which will act as the 'front door' for the electronic version of *Chem. Comm.*, in line with the structure of its printed equivalent. Figure 1 illustrates the overall structure of the conversion process. Using this model, one can develop further tools to convert the archive to other formats, as and when they are needed. This allows the publisher to provide a variety of different versions of the same information, according to the taste of the user, without having to store huge amounts of duplicated data.

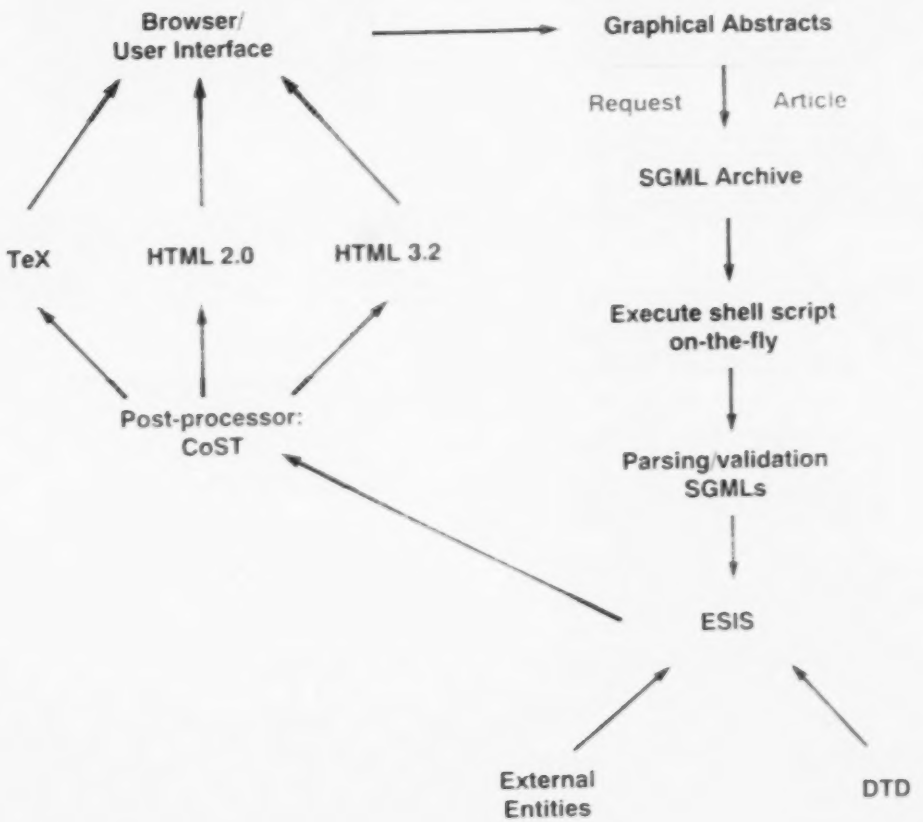

Figure 1: Outline of the document delivery process. The user accesses the Graphical Abstracts page, requests a full article and receives a formatted version via on-the-fly conversion from the SGML archive using the Tcl-based post-processing tool, CoST.

2.3. Added value

An electronic version of an article which is identical in both content and layout to the printed version is an opportunity missed. From the publisher's point of view it is both uneconomical and largely pointless to produce two

BEST COPY AVAILABLE

parallel versions of the same thing. The WWW offers exciting opportunities, in particular in the field of chemistry, to present data in ways which are impossible on the printed page: to 'add value' to the article concerned. Examples of added value could be to embed interactive 3-D molecular viewers, using *Virtual Reality Markup Language* (VRML) or to use a mechanism based on *Multi-purpose Internet Mail Extensions*, or MIME, types to launch external applications automatically when a piece of embedded data is accessed by the user. NMR spectral data, reaction schemes and equations can also be treated similarly. We have called this concept 'hyperactive molecules' and demonstrated it in a number of applications (Refs 12-14). Indeed, one can envisage any object in an article becoming an interactive feature, e.g. equations could be dynamically linked to graphing or symbolic algebra tools. Other possible features of an electronic journal include searching and retrieving data on-the-fly, indexing, 'hyperglossaries', keyword searching and forward referencing. All but the latter are self evident — forward referencing is the concept that a document in the archive can be updated dynamically whenever a subsequent document refers back to it, so generating a hypertextual link 'forward in time'.

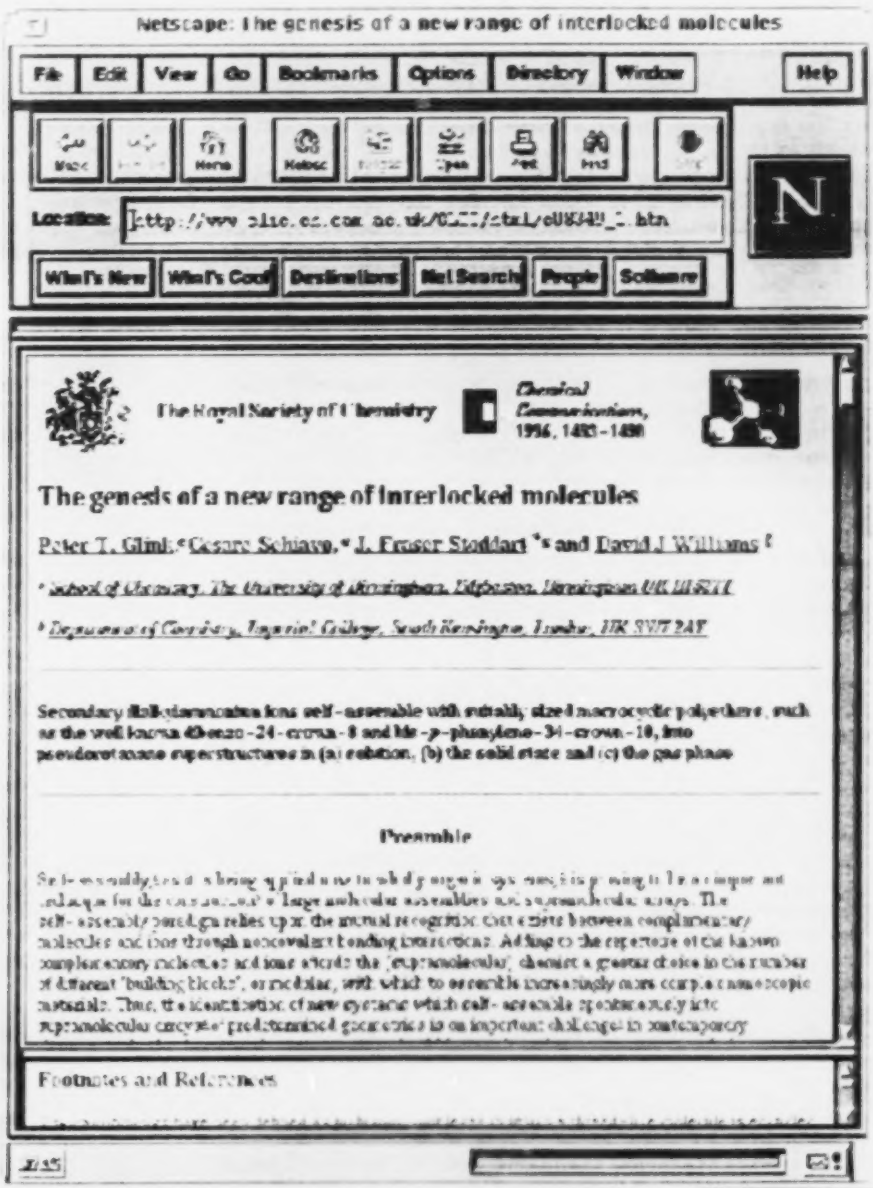

Figure 2: Screen dump illustrating the feature article by Fraser Stoddart and co-workers. The article is displayed within the Netscape Navigator window. Underlined text is hypertextually linked to annotations which appear in the footnote frame. Molecular information in the article encoded in Protein DataBase (PDB) format appears as an animated in-line molecular model.

BEST COPY AVAILABLE

As an example of what can be achieved we have recently presented an enhanced version of a feature article taken from *Chem. Comm.* The paper by Fraser Stoddart and co-workers (Ref 15) describes recent advances in the area of the molecular design of interlocking molecules. In the electronic version figures in the paper version depicting molecular structures become animated 3-dimensional models using the Molecular Design Ltd *Chemscape Chime*™ 'plug-in' for the Netscape browser (Ref 16). This product, based on the RasMol molecular viewer written by Roger Sayle (Ref 17), allows the reader to visualise complex structural information easily. Figure 2 depicts the article as it appears on the reader's screen. Note in particular the use of frames (a non-standard Netscape Navigator feature now included in the specification of HTML 3.2) to provide annotations. Other features in the demonstration article include hypertextual markup of spectral information and extensive cross-referencing. We have also used this technology to provide course notes, tutorials and workshops in a third year undergraduate module offered at Leeds on computer-aided drug design (Ref 18). The potential for the development of distributed courses from several participating universities and for open access to such materials should be obvious.

A relatively new development in browser technology are Java applets — small programs written in the Java programming language (Ref 19). In contrast to CGI scripts which run on the server, these programs are downloaded to the client and executed locally. Therefore they do not depend upon the speed of the connection to run. In some respects similar to the concept of plug-ins the advantage is that the user does not need to download and launch external applications in order to utilise Java, but unlike plug-ins a further advantage is that the Java code is a meta-language and hence platform independent. All that is required is a Java-compliant browser such as Netscape Navigator. Therefore the user needs no technical knowledge of the code and the publishers can incorporate any multimedia, animation or other such added value features which they desire, limited only by their coding ability and imagination. Java applets are relatively new technology: however, their immediate appeal has resulted in an explosion of interest and development, and already there are large numbers of excellent applets freely available on the Web. Our ongoing work with Java will be described at the meeting.

3. Conclusions

We have seen that the technology already exists, and is constantly developing, with which to create interactive documents which provide a multimedia experience for the user. There is no equivalent to this on the printed page, opening up many possibilities not only for the display and dissemination of research but also for teaching. The electronic medium will continue to have an increasing impact on the methods of, and attitudes towards, the publication of scholarly works. There is a certain amount of scepticism levelled at electronic publications because they blur the concept of what constitutes a 'publication'. In essence, however, a properly refereed electronic journal is no different from its paper equivalent. The risk is that until the scientific community at large agrees with this premise, to publish on the Web is to risk your work not being accepted as a *bona fide* publication. With all that this entails regarding research assessment exercises within academic institutions, etc., it is hardly surprising that some scientists may wish to publish on paper first. The level of mistrust can also be attributed in part to the proliferation of low quality information on the WWW:a criticism that can, however, also be levelled at the printed medium, for example some tabloid newspapers.

A properly refereed electronic journal, in particular one which has a respected printed version already in existence, has a good chance of becoming a respectable medium for publication. Whether electronic journals will continue to come out as purely parallel versions of their paper equivalents depends upon the attitudes of authors to this novel publishing medium. This in turn depends on the attitude of the scientific community in general. It is likely, however, that there will come a time when independent electronic journals are established and that the publications in such journals will be considered on a par with those of paper journals. After all, a journal's quality is determined by its refereeing process and content, not the medium by which it is presented.

From a publishers' perspective, the development of an SGML database has several obvious advantages, as outlined previously. The one major drawback is the question of maintaining an archive in the event of a particular publishing house ceasing to exist for whatever reason. There is no obligation on anyone to maintain the archive, and if there is no corresponding paper version, the information could be lost. This lends further weight to the feeling that paper journals could not be fully replaced by their electronic counterparts and that the two will exist in parallel. However it is probable that in the future libraries will redefine their function as we move into an electronic age and take on more of a role as primary archive repositories. At present, while publishers experiment with the Web, much information is provided free of charge. This situation cannot continue as more finance is ploughed into hardware, training of staff and so on and the journals go fully online. Publishers will have to develop a pricing structure. The problem is the level at which to charge. The electronic version of an established paper journal needs to be sufficiently different to allow the product to be sold at a price such that any extra production costs can be recovered. This makes added value crucial; the more extra features, the better chance you have of recovering your costs. In the case of a parallel electronic version of an established journal there is no extra cost for refereeing, but a purely electronic journal must address this issue in its pricing structure. This puts the independent electronic journal at an immediate disadvantage and tends to suggest that the <I>status quo as regards the hierarchy of reputation of journals will be extended into the electronic form. This may cease to be the case, however, if the 'added value' of embedded data finds favour amongst the readership. In this case, and in our view this is a very real possibility, the future for purely electronic journals is bright.

The WWW opens up interesting possibilities for the collection of user feedback. A publisher can automatically

monitor, on a basic level, how many accesses individual articles have, but perhaps more interestingly how individual articles are read by the user. By experimenting with the format of a journal one could ascertain if users tend to look at images before the text, or vice versa, and in what order the article is read. Do readers access the abstract, conclusions and then the body of the article, or each section in order? This information is invaluable for publishers in order to make their product more attractive to the user, whether printed or electronic.

There are issues to be resolved regarding data misuse and the copyright of data. Models for dealing with copyright have not yet fully emerged and legislation to deal with this issue will probably be required. Data misuse is another issue, but abandoning the format in the face of what are in truth teething problems is not a sensible way to proceed. One can assume that, as with the technology, the law will advance, given time. There is little to be gained from waiting for it to catch up: the timescale of developments within the computing community is far shorter than that of the politicians and lawmakers, and the basic issues remain the same in spite of software developments.

Speed of access to the Internet is also a problem, as anyone who has tried to navigate the WWW following transatlantic links on a wet afternoon in the UK will readily testify. That such problems are encountered is evidence of the number of people using the system and it must be remembered how comparatively new this technology is. Precedent tells us that there is a reasonable expectation that such problems will be overcome in time. Apart from the obvious solution of building bigger and faster data highways, caching strategies and data compression techniques are already improving the performance of the WWW although these effects are often masked by the increase in usage. That said, the pace of development of the WWW is a testament to the ease with which the Internet allows information to be passed between individuals. A major criticism of the paper-based publication system is the timescale for work to appear: in some cases a year or more after the work was actually submitted for publication. The Internet, in particular in the form of discussion lists, allows scientists to exchange ideas freely and to keep abreast of rapid advancements, so avoiding the duplication of effort. There is a good chance that the use of the Internet will continue to spiral since there is a self-acceleratory aspect to it. The more people that use the Internet, the faster things develop, the more it has to offer, the more people it attracts and so on. From our experience the usage statistics on our Departmental server continue to show near exponential growth.

In summary, the impact of electronics on the publishing process will be enormous but constructive. The information at our disposal as scientists will be enriched by the advance of technology into this new medium for information dissemination. The fear that the printed journal will somehow be rendered obsolete is unfounded. In the near future all the evidence points to the electronic and paper publishing processes coexisting. In all likelihood they will follow parallel tracks, with the traditional printed form continuing to retain its place in the structure of information provision which exists currently. Finally, as the electronic format becomes more widely accepted as a respectable medium for publication, so publishers and authors will be able — and furthermore would be foolish not — to take full advantage of its unique features.

B.J. Whitaker
School of Chemistry
University of Leeds
Leeds
UK
E-mail: b.j.whitaker@chem.leeds.ac.uk

References

[1] James, D., B.J. Whitaker, C. Hildyard, H.S. Rzepa, O. Casher, J.M. Goodman, D. Riddick and P. Murray-Rust (1995) *New Review of Information Networking*, **1**, 61.

[2] *Report to the Joint Funding Councils' Libraries Review Group* (1993) Chairman: Prof. Sir Brian Follett, HEFCE, Bristol.

[3] Knuth, D.E. (1970) *The TeX Book*, Addison-Wesley.

[4] International Standards Organisation (1986) *ISO 8879: Information Processing — Text and Office Systems Standard Generalized Markup Language (SGML)*, Geneva.

[5] Bryan, M. (1988) *SGML: An Author's guide*, Addison-Wesley.

[6] J. English. CoST is available from *http://trystero.art.com/joe/cost/index.html*

[7] J. Clark. sgmls is available from *ftp://ftp.jclark.com/pub/sgmls/*

[8] A specification for HTML 2 can be found at *http://www.w3.org/pub/WWW/MarkUp/html-spec/html-pubtext.html*

[9] A specification for HTML 3.2 can be found at *http://www.w3.org/pub/WWW/MarkUp/Wilbur/*

[10] John W. Oster II (1993) Electronic Publishing Special Interest Group, ISO 12083.

[11] Bachrach, S., P. Murray-Rust, H.S. Rzepa, and B.J. Whitaker (1996) *Network Science*, March. *http://www.awod.com/netsci/*

[12] Rzepa, H.S., B.J. Whitaker and M.J. Winter (1994) *J. Chem. Soc. Chem. Comm.*, 1907.

[13] Casher, O., G.K. Chandramohan, M.J. Hargreaves, C.J. Leach, P. Murray-Rust, H.S. Rzepa, R. Sayle and B.J. Whitaker (1995) *J. Chem. Soc. Perkin Trans. 2*, 7.

[14] Pudiate, P.B. and B.J. Whitaker, work in progress. *http://cbl.leeds.ac.uk/~caypbp/*

[15] Green, S. and B.J. Whitaker, work in progress. *http://chem.leeds.ac.uk/Project/Teaching/cadd.html*

[16] *The Java Language: A White Paper* (1995) Sun Microsystems, Mountain View, CA.

[17] Glink, P.T., C. Schiavo, J.F. Stoddart and D.J. Williams (1996) *J. Chem. Soc. Chem. Comm.*, 1483. *http://www-clic.ch.cam.ac.uk/CLIC/html/c08349_1.htm*

[18] MDL Information Systems, Inc., San Leandro, CA, USA. See *http://www.mdli.comm/mdlhome.html*

[19] Sayle, R. (1994) *RasMol: A Molecular Visualization System*, Glaxo Group Research, Greenford, Middlesex, UK.